



Archiving — Do You Need It?

First of a Two-Part Series

Analyst: Dianne McAdam

Management Summary

I visited a data center in Boston to find out what they liked (and disliked) about some new technology they had acquired recently. The conversation later turned to a discussion about email. I asked how they managed email. I was told that they backed up the email server to tape every day, and then stacked up the tapes in a dusty corner of the computer room. Then I asked if they archived email. I guess they thought I couldn't hear over the computer room noise, so the answer came back the same - they backed up the email server to tape every night and stacked up the tapes in a corner.

Then I asked them what they planned to do when they needed to retrieve some specific emails. They responded that they would search those all of those backup tapes one by one until they found all of the email messages that they were looking for. I asked how long they thought that would take. The reply was, "a long time and we hope we don't have to do it".

Another Boston company had to find all emails to satisfy litigation brought by a disgruntled former employee. Here are some of the component costs of retrieving those emails on backup tapes.

- *Duration:* 6 months to mount and restore years of backup tapes.
- *Cost for professional services:* \$1,000,000 for contracting outside services to mount and restore years of backup tapes.
- *Having a proper archiving system:* priceless. (Actually, it would have taken them about one day and saved \$1,000,000 in professional services, if they had deployed the right archiving solution.)

That short conversation with the enterprise that stored tapes in a dusty corner of the computer room sums up the problem with some thinking today – that backup and archiving are the same. In fact, **backup and archiving are two different processes with two different end results.** Enterprises that have not built archiving infrastructures are missing opportunities to conserve storage, improve performance, and more intelligently manage one of their most important assets – their data. *Part One* of this series will discuss what an archive is and isn't, and explore why archives are important to all enterprises.

Part Two of this series will discuss what factors must be considered when implementing an archival solution. First, let's look at the difference between a backup and an archive.

IN THIS ISSUE

➤ Backups.....	2
➤ Archives.....	2
➤ Database Archives.....	2
➤ Long-term Retention Archives	3
➤ Do You Need to Archive?.....	3
➤ Benefits of Archiving.....	4
➤ Conclusion	4

Backups

Enterprises have backed up data since the first days of computers. **The purpose of backup has not changed over all these years. Backups are designed to make a point-in-time copy of a volume, LUN, file, or dataset for future use in the event that the volume, LUN, file or dataset is accidentally deleted, corrupted, or becomes unusable.** To understand the difference between backups and archives, let's review some common backup facts.

Backup Facts

- Backups can be created on many different media, such as tape, disk, or optical storage.
- Backups are run on a regular schedule – daily for many applications and several times a day for mission critical applications.
- Multiple copies of backups are kept. For example, a full backup and all succeeding incremental backups will be kept until the next full backup is taken. Several weeks (or months) of backups may be kept, depending on the application. Backups taken after specific events, such as a financial period's close, may be retained for several quarters or years.
- Backups are usually not actively deleted. They are overwritten. For example, an enterprise may have established a policy that states that the last four weeks of full backups be retained. When a new full backup is created in the fifth week, the space for the first full backup can be overwritten.
- Multiple, different backup processes can exist within one enterprise. For example, one backup product can be used to backup data within the main data center. A second backup product is used to back up data in remote offices. And another backup product is used to backup laptops and remote workers' computers.

Bottom Line: Backups are designed to protect data. They are used to restore data when that data is corrupted, deleted, or unusable.

Archives

Archive data is *not* backup data. It is *not* disaster recovery data. It is, simply put,

data that must be kept for long periods of time. It could be a copy of the financial data for the close of the last quarter that needs to be kept for years. It could be the x-ray of a patient's broken leg that should be kept for the life of the patient. It could test results about a new prescription drug that must be submitted to the Food and Drug Administration. It could be emails about contracts. It could be lots of different things – but it is not data that will be used to restore items that get accidentally deleted or corrupted. It is data that needs to be intelligently managed and protected for a specific period of time and then properly deleted (not overwritten), when its retention period has expired. There are two generally classifications of archives – *database archives* and *long-term retention archives*.

Database Archives

Databases grow over time. Take the example of a customer order database for an office supply company. New customers are entered in the database every day. Older customers may continue to buy from this enterprise once every quarter or even once a year. Other customers may not buy anything for five years or more. Nevertheless, this customer information needs to be retained when, and if, the customer initiates a new order.

The database gets larger over time. It takes longer to back up and it takes longer to restore. Indexes grow larger also, and performance may suffer. Software vendors have developed database archiving products, sometimes called *active archiving products*, that allow database administrators to weed out inactive customers and move inactive customer data (along with its associated indexes and table structures) to lower cost storage.

Database Archiving Facts

- Database archiving improves the management of large, growing databases.
- Database archiving is not a form of data protection. It moves inactive data to lower cost storage. Databases must still be backed up on a regular basis to protect against hardware and software malfunctions and database corruption.
- Backups are run daily. However, database archive products typically are not run daily but should be run regularly, such as weekly,

monthly, or quarterly. How often such active archiving products be run? It depends on how fast the database is growing.

- Inactive data can be moved to tape, disk, or optical storage. It can be viewed in its new location or can be restored back to the primary database.

Bottom Line: Database archiving products can reduce storage costs by utilizing lower cost storage and improve performance by slimming down the primary database.

Long-Term Retention Archives

What is long-term retention archive data? It is not backup data. It is not disaster recovery data. It is data that must be kept for a specified period of time. For example, aerospace companies typically have policies to retain airplane designs long after the actual airplanes themselves are built, extending the time until the last plane of a particular design is no longer in active service. This data should be stored by archival software in a separate location from backup data.

Long-term Retention Archiving Facts

- Long-term retention archiving is designed to store data for a specific period of time or until a specific event occurs. Data is actively deleted when the retention period has ended. Some data must be retained until a specific event occurs. For example, X-rays must be kept for several years after the death of a patient, and most financial records are kept for seven years. **Not all data needs to be archived.**
- Backups store *multiple* copies of your data; only *one* copy of the data is generally maintained in an archive. The point is that you should not have to search through all the archives to determine whether or not you have the final version of the data you are looking for.
- Some data must be stored in a form that prevents overwriting, erasing, or altering it during the required retention period.
- Backups and archives may have different *recovery time objectives*, commonly called *RTOs*. RTO is the metric that defines the time required to retrieve/restore data. For example, a customer order database may

have an RTO of one hour; that is, the online order database needs to be restored within one hour to meet service level objectives. However, the RTO for the archived customer data may be much longer. If a corporation is involved in litigation with a former customer and is required to produce supporting documentation, the RTO may be one business day.

- Backups and archives always have different *recovery point objectives (RPOs)*. The RPO refers to the age of the data that is used for restores. A database application, for example, may dictate that it be restored from data that is less than six hours old from the outage. This means that the database has to be backed up at least every six hours. The RPO for archive data, on the other hand, is zero; there is only one final copy of archive data and there are no “lost” updates.

Bottom Line: Backups are about restoring data; archives are about retrieving data.

Do You Need to Archive?

For years, many enterprises have set internal policies for how long its data should be kept. One corporation may dictate that all personnel records are kept for ten years after the person has found other employment. Another organization may set the retention period to only seven years. In many cases, compliance regulations, such as HIPAA and SEC 17a-4, no longer leave retention periods up to the individual enterprises but rather dictate the period of time.

Enterprises that are regulated, such as financial institutions or health care organizations, **must** retain information as dictated by regulations that govern their industry. **These enterprises have no choice – they must implement archiving solutions to ensure that the data is properly retained and can be quickly retrieved when government agencies demand copies of the data.** Failure to do so can result in financial penalties, embarrassing negative publicity, and even jail sentences for executives.

However, there are many enterprises that do not fall under government regulations. Some of these enterprises may feel that implementing an archiving solution is not necessary

or at best low priority on a long list of projects that need to be completed. But that thinking is incorrect and can be very costly to the business.

Let's look at several situations that can be common occurrences for non-regulated businesses.

- A business is sued by a former employee that claims that he was denied (in his opinion) a well-deserved promotion to a high visibility position. This company must be able to reproduce all emails, voice mails, instant messages, and employee records, to prove that a more suitable candidate was found for the position. Instant messages, which contained vital information between hiring manager, the disgruntled employee, and human resources, have not been saved. However, the company suspects that the former employee has printed out and saved some very damaging instant messages that will prove that a lesser qualified person was hired for the coveted position.
- A second company has discovered that their office supply business has been gradually losing business to another competitor that just moved into the area. They need to review all previous orders with customers that have recently jumped ship to the competitor to determine if they need to change their volume pricing or start a marketing campaign to win them back. These invoices were sent via email, and IT just informed the marketing group that it will take months to locate all of the emails on backup tapes.

If these two companies had implemented archival solutions, they easily would have been able to retrieve the information that they needed to contest a lawsuit or boost corporate sales.

Bottom Line: Archiving is not just for enterprises that must retain data as specified by law.

Benefits of Archiving Data

Managing data is a challenge faced by all enterprises today as they struggle to keep pace with the growing amounts of data that they are chartered to protect. So how do archiving solutions fit into to the bigger data manage-

ment picture? To understand role of archiving, we need to take a step back and look at the promise of Information Lifecycle Management (ILM).

Information Lifecycle Management

ILM is not a product but a strategy that protects, manages, and provides continuous access to data as defined by business requirements. Different technologies are deployed to move data to the most cost effective storage while meeting *Service Level Agreements (SLAs)*. The promise of ILM is that enterprises can reduce the cost of ownership while continually meeting business and regulatory requirements. ILM ensures that data will always be properly protected and available.

Archiving must be a critical part of ILM. Archiving is not about writing backups to tape and storing them in a dusty corner of the computer room. Archiving is about storing data for long periods of time and retrieving that data when it is needed in a timely manner. Data may be retrieved to meet regulatory demands. It may be retrieved to meet the demands of internal auditors or legal advisors. Or it may be retrieved to analyze market trends or determine the profiles of the customers that place very large (or very small) orders. Archiving can keep executives out of jail but it can also be a company's competitive edge.

Conclusion

Archiving is a key component of the ILM foundation. It is a repository for data that has been moved from high cost storage to, usually, lower cost storage and can be easily retrieved when required. It should not be a standalone solution but must be an integrated solution that can continue to perform as more and more data is archived over time. **A well-architected archiving solution, like other ILM components, can reduce costs by simplifying management.**

Part Two of this series will evaluate different architectural solutions for archiving, to help enterprises determine which solutions best fit its needs.



About The Clipper Group, Inc.

The Clipper Group, Inc., is an independent consulting firm specializing in acquisition decisions and strategic advice regarding complex, enterprise-class information technologies. Our team of industry professionals averages more than 25 years of real-world experience. A team of staff consultants augments our capabilities, with significant experience across a broad spectrum of applications and environments.

- ***The Clipper Group can be reached at 781-235-0085 and found on the web at www.clipper.com.***

About the Author

Dianne McAdam is Director of Enterprise Information Assurance for the Clipper Group. She brings over three decades of experience as a data center director, educator, technical programmer, systems engineer, and manager for industry-leading vendors. Dianne has held the position of senior analyst at Data Mobility Group and at Illuminata. Before that, she was a technical presentation specialist at EMC's Executive Briefing Center. At Hitachi Data Systems, she served as performance and capacity planning systems engineer and as a systems engineering manager. She also worked at StorageTek as a virtual tape and disk specialist; at Sun Microsystems, as an enterprise storage specialist; and at several large corporations as technical services directors. Dianne earned a Bachelor's and Master's degree in mathematics from Hofstra University in New York.

- ***Reach Dianne McAdam via e-mail at dianne.mcadam@clipper.com or at 781-235-0085 Ext. 212. (Please dial "212" when you hear the automated attendant.)***

Regarding Trademarks and Service Marks

The Clipper Group Navigator, The Clipper Group Explorer, The Clipper Group Observer, The Clipper Group Captain's Log, The Clipper Group Voyager, Clipper Notes, and "clipper.com" are trademarks of The Clipper Group, Inc., and the clipper ship drawings, "Navigating Information Technology Horizons", and "teraproductivity" are service marks of The Clipper Group, Inc. The Clipper Group, Inc., reserves all rights regarding its trademarks and service marks. All other trademarks, etc., belong to their respective owners.

Disclosure

Officers and/or employees of The Clipper Group may own as individuals, directly or indirectly, shares in one or more companies discussed in this bulletin. Company policy prohibits any officer or employee from holding more than one percent of the outstanding shares of any company covered by The Clipper Group. The Clipper Group, Inc., has no such equity holdings.

Regarding the Information in this Issue

The Clipper Group believes the information included in this report to be accurate. Data has been received from a variety of sources, which we believe to be reliable, including manufacturers, distributors, or users of the products discussed herein. The Clipper Group, Inc., cannot be held responsible for any consequential damages resulting from the application of information or opinions contained in this report.